

METHODS & DESIGNS

A nonparametric procedure for evaluation of true and false positives

IRWIN POLLACK,¹ *Mental Health Research Institute, UNIVERSITY OF MICHIGAN, Ann Arbor, Michigan 48104*

A procedure is presented for the evaluation of single-point estimates of true and false positives without strong underlying parametric assumptions. The method is based upon the area operating characteristic and the Green area rule. Estimates of sampling error are also available. The procedure is extended to a strong one-parameter relation between true and false positives.

In a wide range of psychological studies, it is useful to obtain a summary measure of performance in terms of the relation between true positives and false positives. For example, in a recognition memory study, the S may respond "old" to items that have been previously presented ("true positives"), but also may respond "old" to items that have not been previously presented ("false positives"). The same methodology can be extended to the detection of "signals," to the identification of emotional signs, or to any phenomenon where true and false positives can be identified.

The most common summary treatment of the true and false positives is in terms of the strong assumptions of the theory of signal detectability, namely, the assumptions of Thurstone's Case V of underlying normal distribution curves of equal variance. With a given true-positive rate and a given false-positive rate, one can infer the distance between two overlapping normal distributions of equal variance. But, from the single-point estimate, we have no knowledge of whether or not the assumption of underlying normal distributions is valid.

In the best of all possible worlds, each

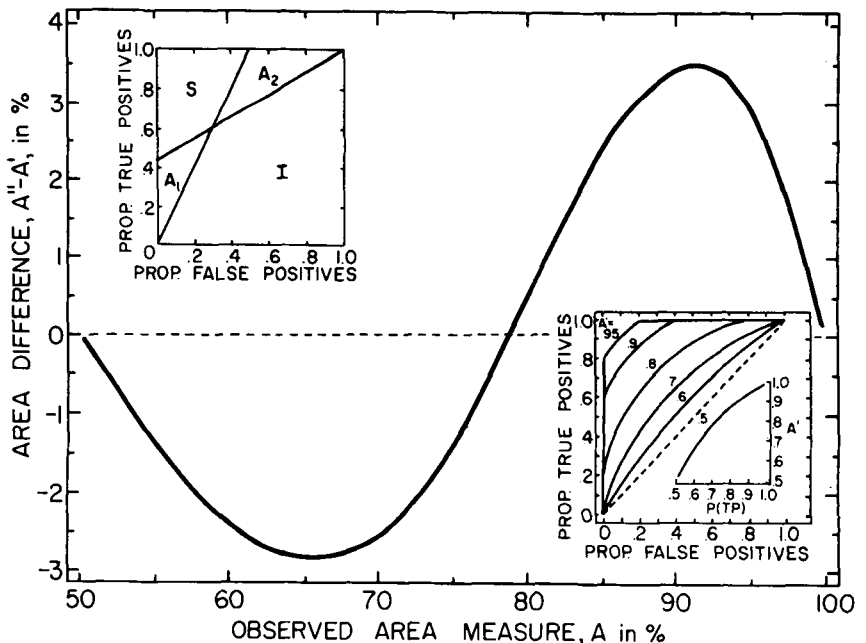
Fig. 1. Green's area measure for area operating characteristics for specified levels of A' . The A' is defined with respect to the upper left insert. Area operating characteristics for specified levels of the A' measure are shown in the right insert. Also drawn, in the lowest right insert, is the relation between A' and the proportion of true positives for (x,y) conditions on the negative diagonal, i.e., where $P(TP) + P(FP) = 1.0$.

experimental condition would yield not just a single observed relation between one true- and one false-positive rate, but rather the whole family of relations. This family defines the *operating characteristic* between true and false positives. With the entire operating characteristic available, there is little need to make strong assumptions about the underlying theoretical distributions, since measures can be obtained directly in terms of the actually obtained operating characteristic. For example, one useful summary measure of performance is the area under the operating characteristic. Green (1964) has shown that this measure predicts the proportion of correct responses in a two-interval forced-choice experiment. The Green equivalence relation does not require strong parametric assumptions about underlying theoretical distributions. The equivalence relation has been verified experimentally in recognition memory by Green and Moses (1966) and in auditory detection tests by Emmerich (1968).

But, in most cases, each experimental condition yields only a single relation between true and false positives. How might we then proceed? We can take advantage of a nonparametric approach to the measure of sensitivity by estimating the

area under a simplified operating characteristic that includes the given point. In the upper left insert of Fig. 1, a single performance is represented at point (x,y), where the x-coordinate defines the false positive rate and the y-coordinate defines the true positive rate. The line between (x,y) and (1,1) represents the performance anticipated if the S performed at level (x,y) on some given fraction of trials and if he arbitrarily responded positively on the remaining fraction of trials, regardless of the experimental conditions. The line between (x,y) and (0,0) represents the performance anticipated if the S performed at level (x,y) on some fraction of trials and if he arbitrarily responded negatively on the remaining fraction of trials. All performances below these two lines are clearly inferior to that represented by (x,y). A similar argument can be developed for the area marked "S," where performance is superior to (x,y). The areas marked "A" are ambiguous in that performance may be either inferior or superior to that represented by (x,y). Pollack and Norman (1964) suggested that we formalize our ignorance by splitting up the ambiguous area to achieve the measure of Eq. 1:

$$A' = 1 + (A_1 + A_2)/2 \quad (1)$$



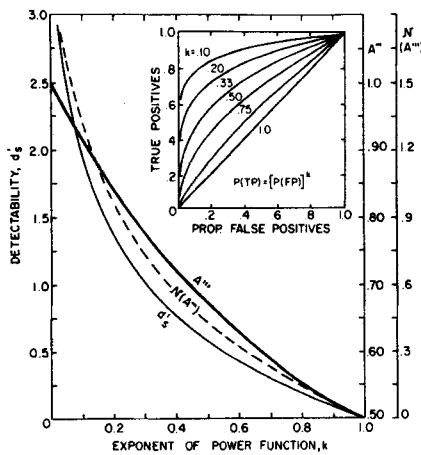


Fig. 2. The relation between k , the exponent of the ratio operating characteristic, and d'_s , a measure derived from the theory of signal detectability (left ordinate). The heavy solid line represents A'' ; the area under the ratio operating characteristic, and scaled on the inside right coordinate. The heavy dashed line represents $N(A'')$; the equivalent normalized value of A'' , and scaled on the outside right coordinate. Ratio operating characteristics for specified levels of k are shown in the upper right insert.

This suggestion leads to an operating characteristic that is symmetrical about the negative diagonal, i.e., about a line between (1,1) and (0.5,0.5). The A' measure will be called the point estimate for area.

Grier (in press) has recently provided Eq. 2 for calculation of A' based upon a single true and false positive rate:

$$A' = \frac{(y-x)(1+y-x)}{4y(1-x)} + \frac{1}{2} \quad (2)$$

where x and y are the false and true positive rates, respectively.

By means of Eq. 2, a range of false and true positive rates can be found that satisfies a constant A' . We shall term the resultant relationship between false positives and true positives as an *area operating characteristic*. All points on the area operating characteristic yield the same point estimate for a given area measure. Examples of area operating characteristics are shown in the lower right insert of Fig. 1. The smaller insert describes the relation between A' and the proportion of true positives for (x,y) conditions which fall on the negative diagonal between (1,1) and (0.5,0.5).

What can we say about the area operating characteristic? In what may appear to the reader (and at times to the

author) as an infinite regress, we can inquire about the area subtended by the area operating characteristic.

An IBM 360 computer was programmed to yield the main results of Fig. 1. In this simulation, for each value of A' , the x -coordinate was varied in steps of .01; the y -coordinate yielded by Eq. 2 was obtained; and the resulting area, A'' , under the area operating characteristic was computed by the trapezoidal rule. The abscissa of Fig. 1 is A' , the point estimate for the area measure. The ordinate is the difference between A'' and A' . A'' is the area subtended by the area operating characteristic for all points associated with a constant A' . The difference between the two measures is 0 at chance (0.50) and at perfect performance (1.00). The difference is negative, i.e., larger A' than A'' scores are obtained for A' values to 0.78 and is positive thereafter. The maximum differences are $-.028$ at $A' = 0.66$ and $+0.035$ at $A' = 0.91$.

Where are we now? We started with an experimental condition that yielded a single estimate of true positives and false positives. We combined these two measures into the A' measure of Eq. 2. The area, A'' , subtended by the area operating characteristic, for defined conditions of the A' measure is obtained from Fig. 1. Finally, we can obtain a measure of the sampling error of A'' , since the sampling error of the area measure is dependent primarily upon its own mean value and is relatively independent of the form of the underlying distributions (Pollack & Hsieh, 1969, Fig. 7).

In partial summary, a nonparametric procedure for the evaluation of a given point of false and true positives, which does not make strong parametric assumptions of specific underlying distributions, is:

(1) From Eq. 2, due to Grier (in press), determine A' . This defines a single point on the area operating characteristic.

(2) From Fig. 1, determine A'' . This defines the area under the associated area operating characteristic.

(3) From the reference listed in the previous paragraph, determine the sampling error of A'' .

Finally, consider the application of the area operating characteristic to one specific strong relation between true and false positives. A simple one-parametric description of operating characteristics has been provided by Egan, Greenberg, and Schulman (1961). This relationship was found to be an excellent descriptor of situations without defined trials, as in "vigilance" studies. Eq. 3 describes their

power relation between true positives and false positives. This relation will be called the ratio operating characteristic.

$$P(TP) = [P(FP)]^k \quad \text{where } 0 \leq k \leq 1 \quad (3)$$

• Sample ratio operating characteristics are presented in the upper right insert of Fig. 2. Note the strong asymmetry of the functions in contrast with those of the lower right insert of Fig. 1.

The area under the ratio operating characteristic, A''' , is simply $1/(k+1)$. The area measure is presented, on the inner right ordinate, as a function of the exponent of Eq. 3. Also presented in Fig. 2, on the left ordinate, is d'_s , a measure derived from the intersection of the ratio operating characteristic and the negative diagonal and then transformed into standard or z -scores, based upon the normal probability curve. Better agreement between the two measures is obtained when the area measure is also transformed to equivalent z -score form, as represented by the extreme right ordinate, $N(A''')$. The suggestions previously directed toward the sampling error of the area measure of the area operating characteristic, A'' , may also be directed toward the area measure of the ratio operating characteristic, A''' .

REFERENCES

- EGAN, J. P., GREENBERG, G. Z., & SCHULMAN, A. I. Operating characteristics, signal detectability, and the method of free response. *Journal of the Acoustical Society of America*, 1961, 33, 993-1007.
- EMMERICH, D. S. Receiver-operating characteristics determined under several interaural conditions of listening. *Journal of the Acoustical Society of America*, 1968, 43, 298-307.
- GREEN, D. M. General prediction relating yes-no and forced choice results. *Journal of the Acoustical Society of America*, 1964, 36, 1042(A).
- GREEN, D. M., & MOSES, F. L. On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, 1966, 66, 228-234.
- GRIER, J. B. Nonparametric indices for sensitivity and bias: Computing formulas. *Psychological Bulletin*, in press.
- POLLACK, I., & HSIEH, R. Sampling variability of the area under the ROC-curve and of d'_e . *Psychological Bulletin*, 1969, 71, 161-173.
- POLLACK, I., & NORMAN, D. A. A nonparametric analysis of recognition experiments. *Psychonomic Science*, 1964, 1, 125-126.

NOTE

1. This research was supported in part by Grant GB 6148 of the National Science Foundation. Mr. Robert Hsieh provided the program for Fig. 1. Although Eq. 2 was required for the earlier development of area operating characteristics, its explicit statement is due to J. B. Grier.